(E74-10491) AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT Quarterly Progress Report (California Univ.) 14 p HC $4.00 CSCL 02C
N74-22015
Unclas
G3/13 00491

SPACE SCIENCES LABORATORY

E7.4-10491.
CR-136892

AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT

A report of work done by scientists at the Center for Remote Sensing Research, University of California, Berkeley, under NASA Contract No. NAS 2-7567

Quarterly Progress Report
28 February 1974

UNIVERSITY OF CALIFORNIA, BERKELEY

AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT
(EPN NO. 382)

Principal Investigator: Robert N. Colwell
Co-investigator: Andrew S. Benson
Contributors: Catherine E. Brown
Claire M. Hay, Nancy A. Jones
Remote Sensing Research Program
Space Sciences Laboratory
University of California
Berkeley, California 94720

1.0 INTRODUCTION

During the past quarter, the emphasis of the Skylab agricultural investigations has concentrated on the quantitative evaluation of the S190A and B photographic data from the Earth Resources Experimental Package (EREP). Analysis of data from the S192 multispectral scanner is being delayed due to the lack of noise-free computer compatible tapes. The following sections of this quarterly progress report document those tasks which have been accomplished during this reporting period, and those tasks which are projected for the remainder of the contract performance period.

2.0 WORK PERFORMED DURING THE PERIOD COVERED BY THIS REPORT

2.1 Status of Skylab Data and High Altitude Aircraft Imagery

2.1.1 Skylab 2

All photography of our study areas that was acquired from Skylab 2 (Track 63, 3 June 1973) has been received along with its supporting high altitude aircraft imagery. During the June overpass, the S190B camera was not in operation; however, high quality S190A photography was acquired.

To date, only one S192 digital tape (containing 5 seconds of data) has been received which covers the southern portion of the San Joaquin Valley test site (see Figure 1). Of the ten channels on this tape, four were considered unusable -- channels 1 and 13 were almost completely saturated, and a high degree of systematic noise was present in channels 6 and 11. The remaining channels -- 2, 3, 8, 9, 10 and 12 -- were usable for CALSCAN* analysis. Complete coverage of the test area with

*CALSCAN is the RSRP version of the LARS-Purdue pattern recognition program adapted to the CDC 6600/7600 system at the University of California, Berkeley.

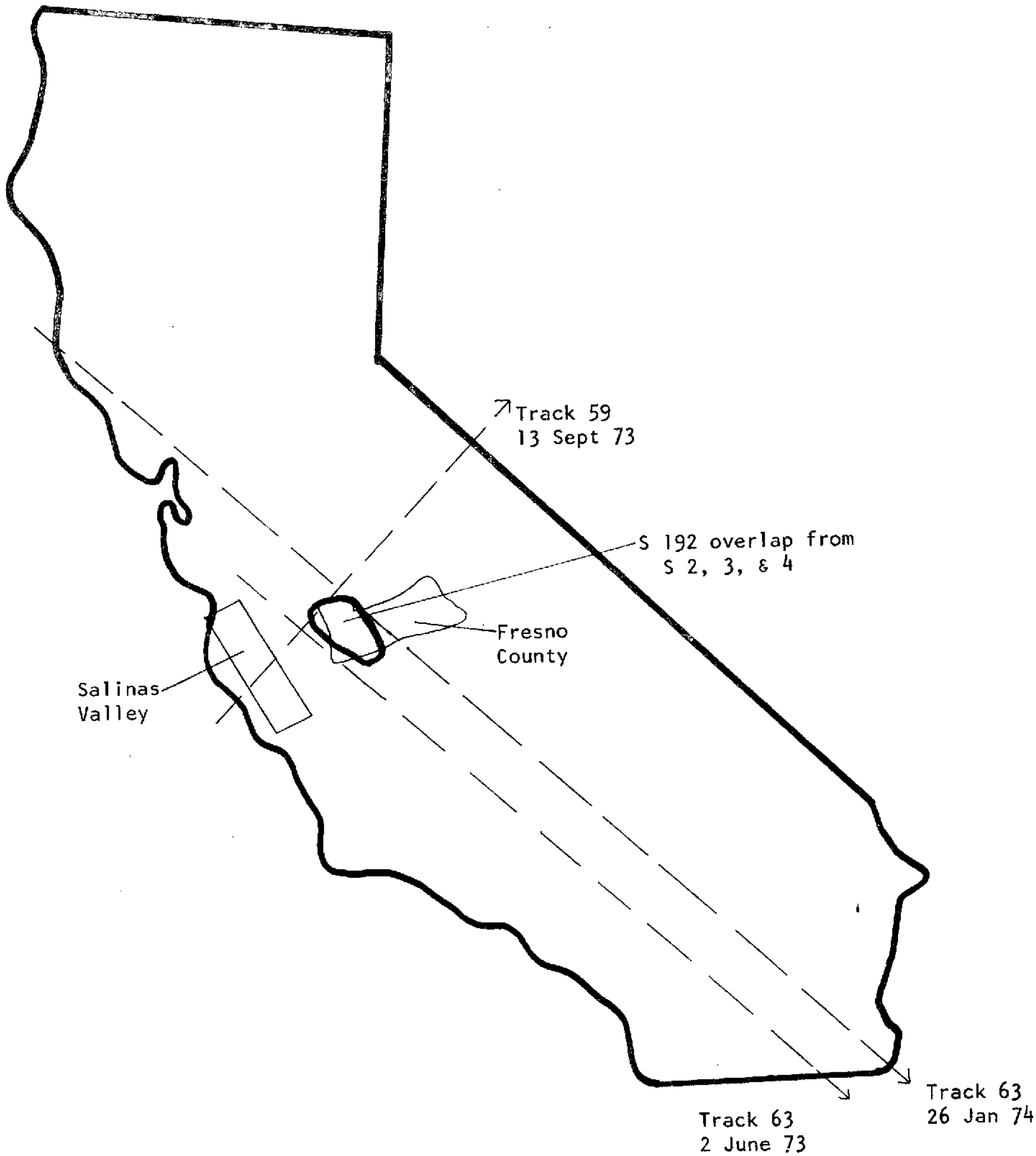


Figure 1. The above map shows the state of California and the location of the agricultural test sites with respect to the ground tracks from the three Skylab missions.

noise-free data has been ordered and is scheduled to be received in April 1974.

2.1.2 Skylab 3

One EREP pass along Track 59, 13 September 1973, was made during Skylab 3. Color and color infrared imagery from the S190A camera has been received along with color imagery from the S190B camera. Noise corrected data from the S192 scanner has been ordered and should be received in June 1974.

2.1.3 Skylab 4

Two EREP passes were made along Track 63 on 3 December 1973 and on 26 January 1974. Data deliveries for S190A and S190B cameras are expected in late spring and early summer; no delivery date has been given for the S192 scanner data. Because the Skylab vehicle returned to its original orbit during this mission, the amount of overlap coverage with preceeding missions is limited, particularly for the S192 multispectral scanner. This overlap area is shown in Figure 1.

2.2 Manual Crop Inventory

To date, most of the identification and inventory of agricultural crops using satellite data (i.e., ERTS-1) have been done semi-automatically by computer analysis of digital tapes with human inputs of training materials. Although such systems have achieved very accurate results in certain test areas, there is still a need to develop techniques for manual interpretation of satellite data for agricultural resource inventories. While computer-based systems may ultimately provide the most efficient method for gathering agricultural statistics of extensive areas, at the present time the human interpreter represents the most expedient way to perform an operational inventory in the United States. For many of the emerging nations of the world where both national agricultural statistics and computer systems may be non-existent, the data gathered by human interpreters can provide a valuable input to the management decisions for agricultural resources.

In any attempt to develop efficient techniques for the human interpretation of Skylab data, several factors must be considered: (1) because of the large areal coverage of the imagery, 100 percent image interpretation of the entire frame for detailed information is not practical, (2) a simple method is needed to evaluate the accuracy of the interpreter's estimates and, if necessary, to adjust these estimates, and (3) the low resolution of the satellite imagery makes accurate acreage estimates by human interpreters impractical.

An inventory technique employing a double sample utilizing point data is being designed to deal effectively with these constraints. A single crop, namely cotton, has been selected on which these techniques

are being tested. For the first stage in the sample design (large sample), the interpreter will determine the presence or absence of cotton at a large number of points located throughout southwestern Fresno County. (This test area was chosen because of the availability of sequential photographic coverage from SL-2 and SL-3). These data will then be used to estimate by means of manual interpretation the proportion of the area that is planted to cotton. This proportion, when multiplied by the total area being inventoried will give an unbiased estimate of the total crop acreage. The second stage in the inventory will consist of a subsample of the large sample. This subsample will be drawn from those points from the large sample which fall within "ground truth" fields for which the crop type is known. The correlation between "ground truth" and image interpretation estimates will be used to evaluate interpreter accuracy and to calculate a ratio estimator to adjust the interpreter's estimated proportion.

It should be noted that in operational surveys which would utilize real time satellite data, the method for selecting the subsamples would differ slightly. The ground samples would not be chosen until after the satellite data had been received and interpreted, thus eliminating the need for maintaining permanent ground data points. However, the technique as first described still has potential applications since many governmental agencies maintain some form of permanent ground plots for which accurate data are collected. This technique would allow the agencies to expand these data more accurately with the use of remotely sensed data.

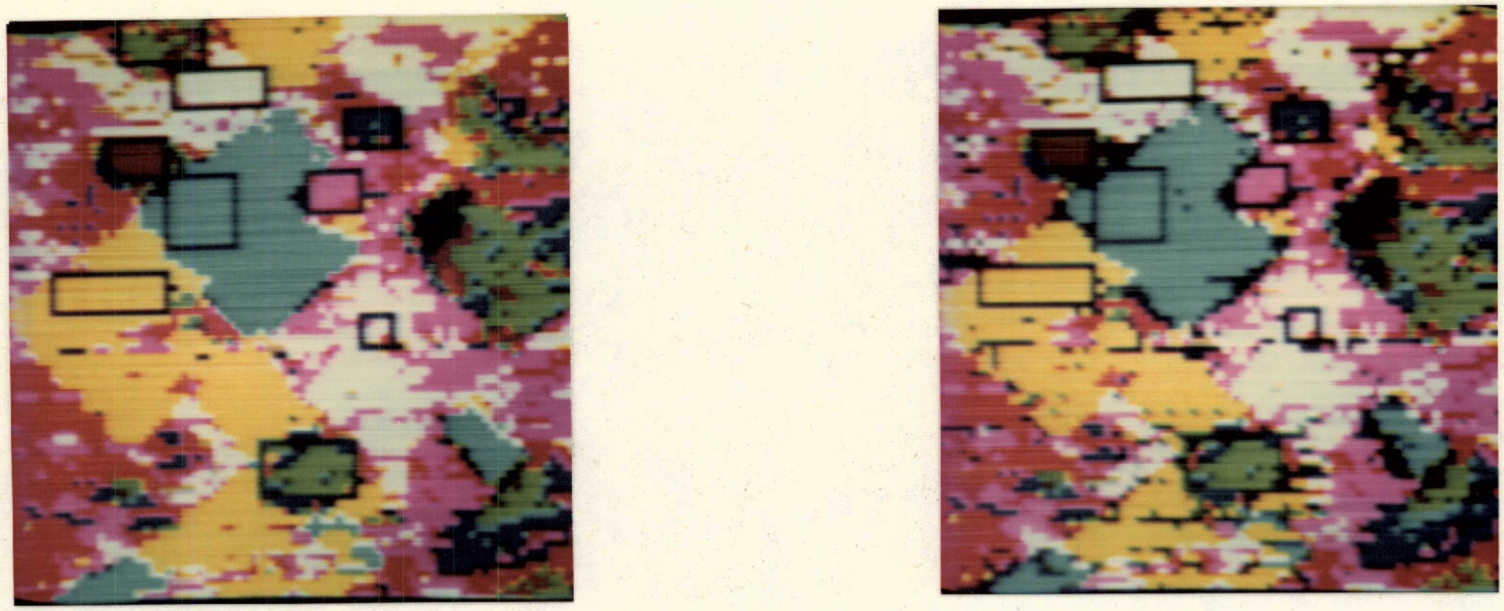
2.3 Automatic Interpretation

The planned automatic data processing studies call for extracting agricultural information from EREP computer-compatible tapes and photographic transparencies. The processing of both types of data will give a comparison of classification results using an analog storage medium (film) versus digital storage (tape).

2.3.1 Multispectral Scanner Data

Initial experimentation with data from the S192 multispectral scanner was begun using only six channels: 2, 3, 8, 9, 10 and 12 (see Section 2.1.1). Spectral data for seven crop types, viz., immature barley, mature barley, cucumber, plowed-fallow, tomato, cotton and alfalfa, were extracted from a section of the digital tape. From a statistical analysis of these data it was determined that channels 2, 3, 8 and 12 only would be needed for the CALSCAN classifier because the data contained on channels 9 and 10 were highly correlated with those on channel 8 and therefore were considered redundant.

Of the 3,500 acres (1,400 hectares) of cropland that were classified, 75 percent of the fields were identified correctly. The visual results of a portion of this area are shown in Figures 2a and 2b.



(a) | 1 mile | (b)

yellow - mature barley
red - tomato
white - cotton
orange - fallow
pink - potato

purple - cucumbers
aqua - barley stubble
blue - safflower
green - alfalfa
black - everything else

Figure 2. The results of CALSCAN analysis of the S192 scanner data using a medium (a) and a fine (b) threshold. The area shown here is located west of Fresno, California. Data within the black rectangular boxes were used for training the automatic classifier. See text for additional discussion.

Figure 2a was formed from a CALSCAN analysis which used a medium threshold decision rule and Figure 2b used a fine threshold. It is evident that even when using the medium threshold, a considerable amount of noise, as indicated by horizontal black lines, is still present on the tapes. In addition, extensive classification errors occurred, indicated by other color anomalies within fields, which was not due to the quality of the data per se, but to the immaturity of the crops at the time of the Skylab overpass which makes crop discrimination difficult. An example of this confusion is shown in both Figures 2a and 2b for immature cotton (white) and immature cucumbers (purple). Despite the noise present on the data and the non-optimal acquisition time of the data, these results are very encouraging when one considers that neither channel 4 (green) nor channel 6 (red) was available for use in the classifier. Once these two channels become available, we expect the classification accuracy to improve considerably.

2.3.2 Multispectral Camera Data

All four bands of the S192 black-and-white photographic data obtained during the SL-2 mission were scanned with a microdensitometer for an area covering the northern half of the Salinas Valley test site. The scan interval along both the X and Y axes was .01 inches, and the aperture of the scanner was set at .001 inch diameter so that each data point represented a 1.12 acre spot size on the gorund. The resulting density measurements were recorded on magnetic tape and were subsequently reformatted on the CDC 6600/7600 system so that they would be compatible with the CALSCAN program. Training fields were extracted from the reformatted tape, and the STAT subprogram* of CALSCAN was used to produce various descriptive statistics.

The mean density and standard deviation as scanned from the four bands for tomato, asparagus, sugar beets, lettuce, carrots, cauliflower and beans are shown in Table 1; these spectral densities, as shown in Table 2, overlap in every band. It is evident from these statistics that data from one band of imagery alone will be insufficient to discriminate between these crops. Uniform appearance, i.e., small deviation from the mean, facilitates automatic classification. However, the crop types which display these small deviations, such as tomato and asparagus on the .8 - .9 μm band, also show a correspondingly small difference between their actual means, and thus they have virtually identical density distributions. Accurate classification of these vegetable crops using either automatic or manual techniques will be difficult because of the similar states of development of these crops

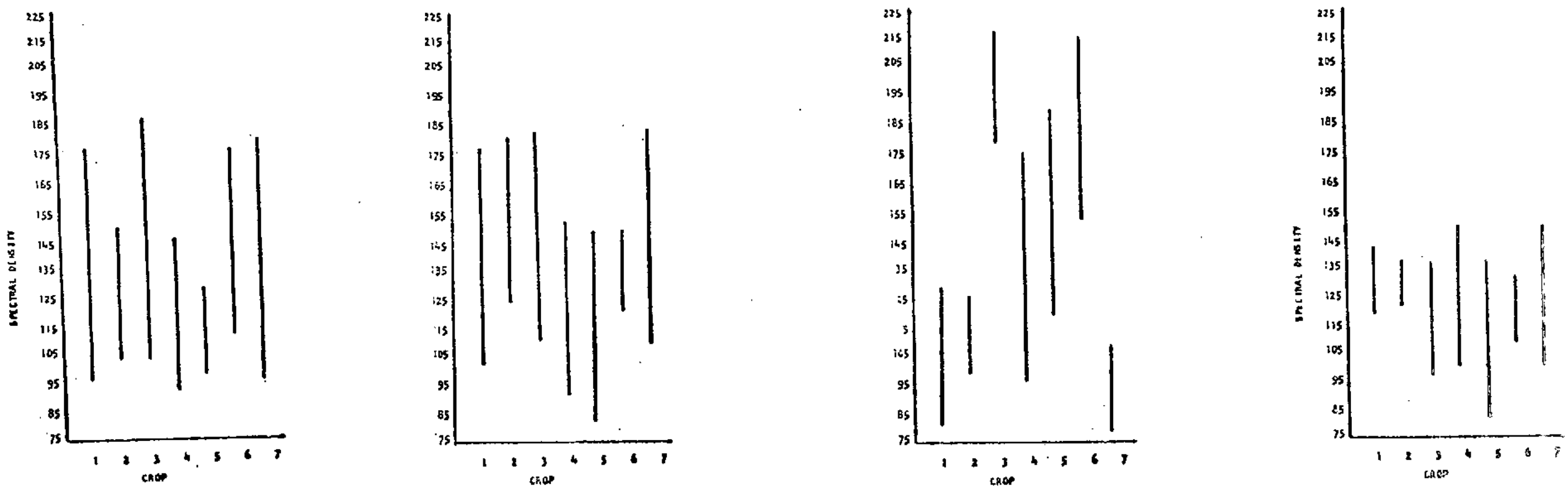
*The STAT subprogram of CALSCAN performs statistical analyses on specified training fields and prints the results in various forms as directed by the user. These data indicate which classes are separable on the basis on training data given and also which features offer the best separation.

TABLE 1. STATISTICS BY CROP TYPE
DERIVED FROM SCANNED
S190A PHOTOGRAPHIC DATA

		BAND 1 .7 - .8 μm	BAND 2 .8 - .9 μm	BAND 5 .6 - .7 μm	BAND 6 .5 - .6 μm
TOMATO	$\overline{X}$	97.96	129.13	127.99	136.22
	S.D.	28.91	11.97	36.09	40.13
ASPARAGUS	$\overline{X}$	109.08	126.31	151.74	126.64
	S.D.	13.65	7.38	28.56	24.01
SUGAR BEETS	$\overline{X}$	196.92	115.47	146.36	145.41
	S.D.	21.00	19.96	37.67	44.37
LETTUCE	$\overline{X}$	131.48	122.50	121.47	119.17
	S.D.	39.23	23.37	30.40	27.31
CARROTS	$\overline{X}$	154.30	108.77	112.92	114.34
	S.D.	33.24	27.24	33.38	13.93
CAULIFLOWER	$\overline{X}$	184.50	118.29	145.71	144.06
	S.D.	31.03	10.42	23.15	32.19
BEANS	$\overline{X}$	90.05	122.56	145.32	137.82
	S.D.	20.88	23.79	37.65	42.01

Note: Shown here are the mean density ($\overline{X}$) and standard deviation for the seven crop types present in the Salinas Valley as scanned from the four black-and-white EREP transparencies. These statistics will be used in part to select the optimum bands for automatic classification of crop type using the CALSCAN program.

TABLE 2. DENSITY DISTRIBUTIONS BY CROP TYPE
DERIVED FROM SCANNED S190A PHOTOGRAPHIC DATA



.5 - .6 µm　　.6 - .7 µm　　.7 - .8 µm　　.8 - .9 µm

CROPS
1 = Tomato
2 = Asparagus
3 = Sugar Beets
4 = Lettuce
5 = Carrots
6 = Cauliflower
7 = Beans

Note: The relative distribution of densities of these vegetable crops on all four S190A black-and-white transparencies taken in June 1973, indicate that (1) no one band alone will be sufficient to discriminate between the crops, and (2) even by using the data from all four bands, accurate crop discrimination may not be possible. See text for further discussion.

at the time of image acquisition. In June in the Salinas Valley, tomato, carrots, lettuce and cauliflower were just emerging from the soil and had the same basic appearance as bare soil. Improvement in crop identification accuracy is expected from the SL-3 data because all crops will be fully mature. When SL-2 (June) and SL-3 (September) data taken of the same area are combined, variations in cropping cycles will greatly facilitate discrimination between crops that have similar spectral reflectance characteristics on any single date.

3.0 WORK PLANNED FOR NEXT REPORTING PERIOD

3.1 Introduction

Due to the delays that are expected in receiving all S192 data it is doubtful whether much additional work with these data will be accomplished during the final quarter of the contract performance period. However, complete analysis of all photographic products will be completed within this time frame.

3.2 Manual Interpretation

3.2.1 Crop Inventory

The double sample design as described in Section 2.2 will be completed. The final results from this study will include acreage estimates, sampling error, and an overall evaluation of the technique.

3.2.2 Irrigated Land Study

There presently exists an important need on the part of the Department of Water Resources of the State of California for a periodic tabulation in any given year of the statewide acreage of agricultural land receiving irrigation. A study will be performed to investigate the extent to which this tabulation can be accomplished using two dates of EREP data and the appropriate sampling designs. This study will not be able to give a valid yearly statistic of the irrigated lands within the test area due to the fact that only two dates of imagery are expected to be available and imagery taken at other dates would be required. However, it will help determine (1) the feasibility of making such inventories, (2) the expected accuracies of such inventories, (3) whether similar techniques could be used on ERTS-1 imagery, and (4) the probable cost for making such inventories on a statewide basis.

3.3 Automatic Interpretation

3.3.1 Multispectral Scanner Data

As stated in previous sections of this report, analysis of the S192 data has been delayed. The extent to which this analysis can be

completed by the end of the contract performance period will be contingent upon the timely receipt of noise-corrected data for all channels covering the entire test area from SL-2, -3 and -4. With these data we will be able to (1) determine the relative contribution each channel makes in the CALSCAN classifier, (2) modify existing software to enable the overlaying of ground Tracks 63 (SL-2), 59 (SL-3) and 63 (SL-4) to facilitate multidate analysis, and (3) compare the spatial and spectral resolutions of Skylab and ERTS with respect to accuracy of crop identification and costs of data processing.

3.3.2 Multispectral Camera Data

Black-and-white S190A transparencies from SL-3 have just recently been scanned, and the data are now being processed. We expect that the classification accuracy achieved with these data will be much more accurate than that from the SL-2 data, because the crops within the test area have matured and should have more unique spectral signatures than they displayed in the June pass of SL-2.

3.4 Multistage Sampling of Agricultural Resources

The culmination of our Skylab agricultural experiments will be the demonstration of agricultural survey techniques using a multistage sampling model. Through use of the discriminant analysis techniques described below, Skylab photo and scanner data will be combined with aerial photos and ground data to give an estimate of crop acreage within the areas common to Track 63 of Skylab 2 and Track 59 of Skylab 3. We have not yet chosen which crop or crops will be inventoried; this will be determined after the data from the S192 multispectral scanner have been analyzed. Under normal operational conditions, the optimum times for obtaining remote sensing data for a successful agricultural inventory are known and are carefully adhered to. However, since this is not possible under the restrictions of the EREP missions, the crops to be surveyed will be those which can best be inventoried with the available Skylab data.

The multistage model relies heavily upon the first stage in which the information extracted from the Skylab data by human interpreters and automatic classifiers provides the initial estimates of the resource. The first step of the data extraction process will be one in which human interpreters stratify all fields within the area of interest into broad land use categories and crop classes based on their appearance on the Skylab S190A ektachrome imagery. At this time political and geographic boundaries will be superimposed on the imagery to further define the areas of interest. Next, a number of fields which represent the various agricultural resources of interest will be

selected from each stratum to train the discriminant analysis program. The identities of these fields will be determined from ground data and/or the interpretation of aerial photos. The number of training fields required for each crop class will be dependent upon the variability of the spectral signatures of the crops present. This variability is caused by such factors as different cropping practices, local soil differences, and genetic variations within each particular crop type. For example, a crop such as alfalfa which may be in several stages of maturity throughout the survey area at the time of image acquisition may require five or more fields per stratum for adequate training, whereas only one training field per stratum may be needed for a less complex crop type such as corn. After the fields have been chosen, they will be located on, and extracted from, the spacecraft imagery. The multispectral data from the training fields will be run through the discriminant analysis program to obtain a point-by-point classification of the entire area by strata. This procedure will provide an initial estimate of the acreage of the vegetation classes.

In the second stage of the model, the results of the discriminant analysis will be sampled to determine their relationship to ground estimates of the resource. Sampling units (SU) will be defined by dividing the entire area into rectangular blocks. The size and shape of these blocks will be determined by (1) the information requirements, (2) the variability that occurs in the SU estimates as the block size changes, (3) the costs of further subsampling, and (4) the resolution of the scanner imagery.

Probability sampling is expected to provide the most efficient sampling design that can be applied in the second stage of the model. Probability sampling is a special case of the mean of the ratios estimation where samples are allocated proportional to the expected variance of the X_i estimate. For this model, the total value of the i^{th} SU, denoted by X_i, is evaluated by

$$X_i = \sum_{m=1}^{M} \sum_{j=1}^{J} I_m V_j$$

where $\begin{cases} I_m = 1 \text{ if } C_m = j \\ I_m = 0 \text{ otherwise} \end{cases}$

C_m = crop class for the n^{th} "pixel" (picture element) of the SU, as determined by the discriminant analysis,

M = the number of "pixels" per SU,

V_j = the crop class being evaluated

J = the number of crop classes.

The value or weight (v_j) is assigned to rank the various crops or vegetation types based on their relative importance to the survey. In an agricultural inventory where total dollar value is the objective, the v_j's are the average dollar values per "pixel" of the various crops (j). If the making of an inventory of a single crop is the objective, the value of the crop of interest is 1 and all other v_j's are set to zero. In many cases this weighting factor is primarily affected by the marketing conditions of each crop, and is highest for those crops for which acreage estimate errors are most important.

The variance of the population is estimated by

$$S^2 = \frac{1}{N-1}\sum(X_i-\bar{X})^2$$

The number (n) of SU's to be selected for photo and ground measurement when no remote sensing information is available is determined by

$$n = \frac{Nt^2S^2}{N(AE)^2 + t^2S^2}$$

where AE = the allowable error, in units of value

t is a value obtained from "students t" tables and

S^2 is as defined previously.

The n points are then selected from the list of SU's proportional to their estimated value.

The selected SU's are then carefully transferred to the corresponding high flight photography where precise field size measurements are made for use later in adjusting the acreage estimates obtained from the classifier.

From high flight images, low altitudes images, ground identification and historical data, the "correct" classification for each field in the SU is determined, down to crop type and maturity.

The total value for the area ($\hat{T}$) is estimated using the probability of selection (P_i) and the photo/ground estimate of SU value (Y_i) by means of the relation,

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{P_i}$$

where $$P_i = \frac{x_i}{\sum_{i=1}^{N} x_i}$$

The variance of the estimate for $\hat{T}$ is

$$s_{\hat{T}}^2 = Var(\hat{T}) = \frac{1}{n} \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - \hat{T}\right)^2$$

If the photo/ground estimate (Y_i) were to be perfectly proportional to the remote sensing estimate (X_i), only one ground sample would be needed to determine the proportionality constant. More realistically however, the number of ground samples (n) for future surveys is estimated by:

$$n = \frac{Nt^2 s_{\hat{T}}^2}{N(AE)^2 + t^2 s_{\hat{T}}^2}$$

This probability sampling model is appropriate when a single parameter such as "acreage of a single crop", "value of all the crops present", or "demand for irrigation water" is desired. Such a model can be replaced by a regression sampling model if estimates on a crop-by-crop basis are required; however, the regression sampling model will only meet the allowable error criterion for the total value of all crops by strata.